A SEMI-AUTOMATED WORKFLOW FOR PRODUCING TIME-ALIGNED INTERMEDIATE TONAL REPRESENTATIONS

LAURA MCPHERSON AND EMILY GRABOWSKI
(DARTMOUTH COLLEGE)

ICLDC 9, March 2, 2017

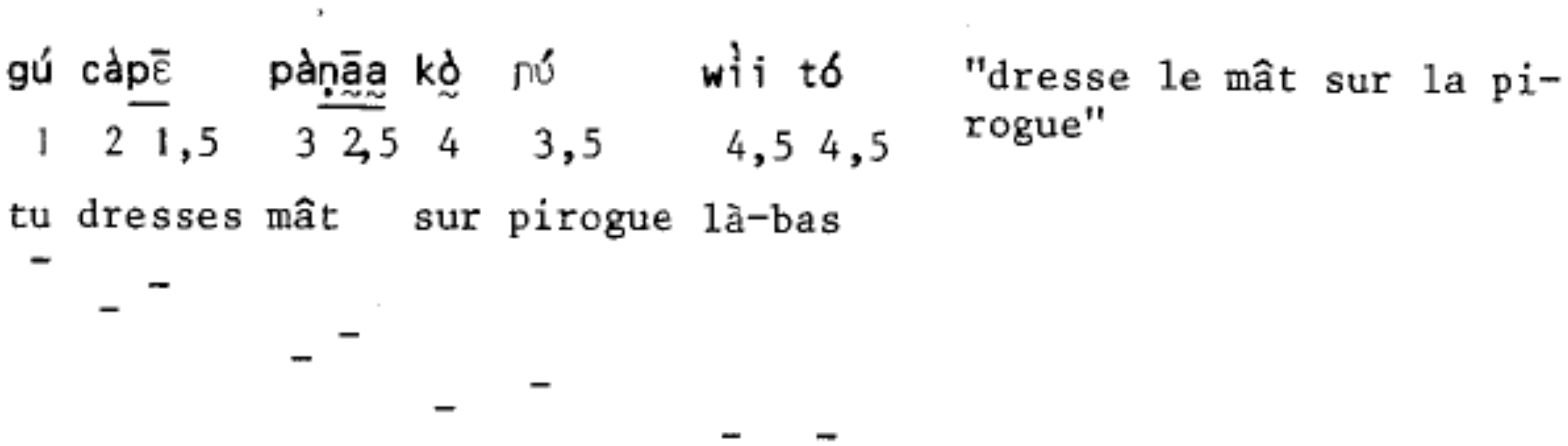
Introduction

The problem

- Tone is notoriously difficult
  - Inherently relative
- Early transcriptions unreliable
  - How many contrastive levels?
  - Are contours phonetic or phonological?
- Researchers are not always trained in tone
  - Community members
  - Linguists too

Annotations

- “Phonetic” annotations can be unsystematic and difficult to digitize:
  - [¯ –] [– _] [_ /] etc.



gú càpɛ̄ pàṇā̰a̰ kò̰ ɲú wìi tó "dresse le mât sur la pirogue"

1 2 1,5 3 2,5 4 3,5 4,5 4,5

tu dresses mât sur pirogue là-bas

Annotation of Numèè by Jean-Claude Rivierre (1973:134)

Annotations

- Phonological analyses often abstract, obscuring phonetic underpinnings

ɲḭ́ á mwḭ̀ḭ b̲ɛ̲́ɲḛ́ṛḛ́-ṛá "il est plus fort que moi"

il qui fort dépasser-moi

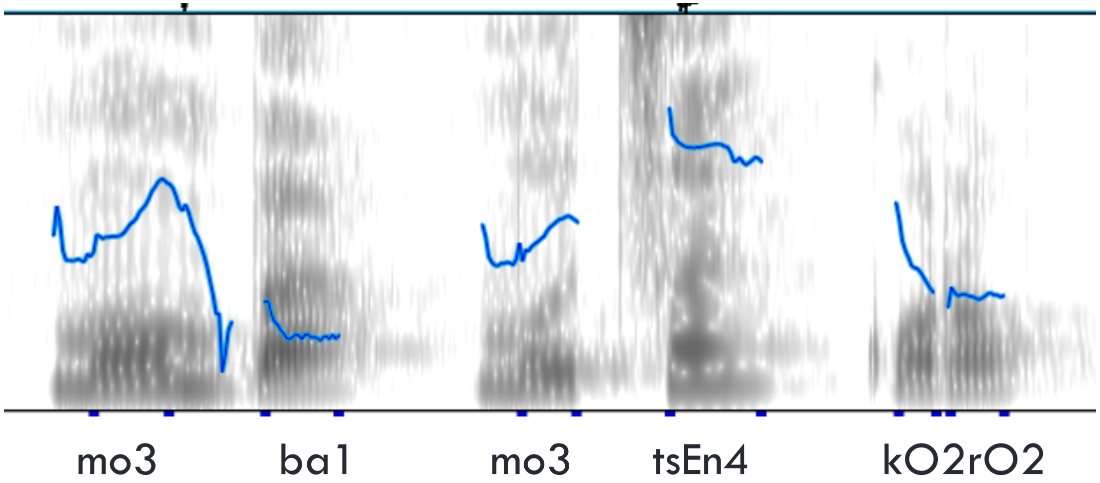
- No guarantee the researcher’s analysis is correct

The proposal

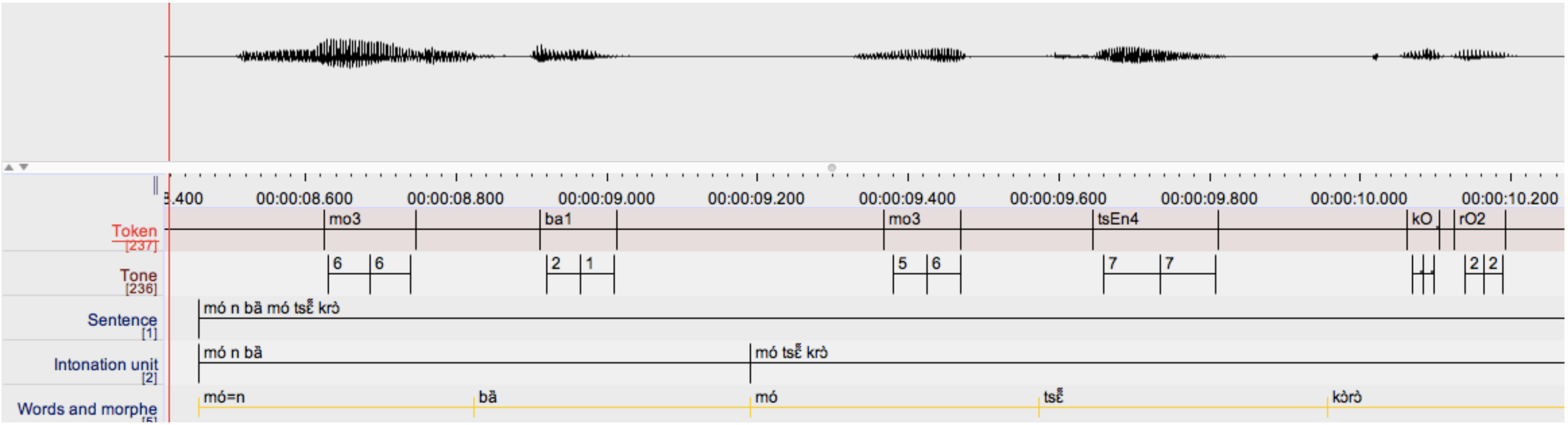
- We need a tool to help produce **objective, replicable** tone annotations from Day 1
- Desiderata:
  - Based on acoustic data (f0)
  - User friendly
  - Easily interpretable annotations
  - Interface with existing software and technology

The proposal

We want to go from this (f0, messy, hard to interpret):



To this (discrete levels, able to be digitized and included in annotations):



ATLAS

- ATLAS: Automated Tone Level Annotation System
- Annotates recordings for phonetic tone level based on normalized f0
  - Output is time-stamped
  - Can be imported into ELAN as a tonal tier
- NB: Does not replace the need for phonological analysis, but can appear alongside
- Analytical upshots:
  - Produces a searchable corpus of tone tied to other grammatical information
  - Annotations can be used to study phonetics or intonational realizations of tone

Today’s talk

□ Existing technologies for tone

□ Overview of ATLAS

□ Research applications

□ Conclusions/future work

Existing technologies for tone

Existing technologies

- Tone is relatively underserved technologically, but a few tools have been developed
- Focus on analysis of lexical/phonological tone, not surface/phonetic representation
- Two broad categories:
  - Hidden Markov models (language specific)
  - Clustering (language independent)

Hidden Markov Models

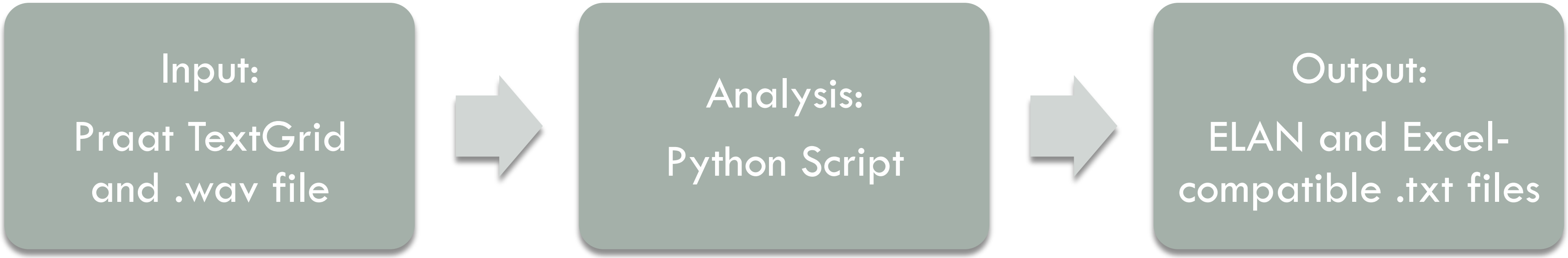
- Hidden Markov Models
  - Mandarin Chinese (Wu, Zahorian, and Hu 2013, Yang et al 1984)
  - Cantonese (Tan Lee et al. 1995)
  - Thai (Cooper-Leavitt 2016)
- Tone requires more context than most HMMs utilize (Bird 1994)
- Tools are limited to a handful of well-studied languages
  - All of which are East Asian contour-based tone systems

Clustering

- Not many computational tools for unstudied tone systems
- Toney (Bird and Lee, 2014)
  - Displays F0 contour on a canvas and allows the user to group similar contours together
  - Does not appear to be in active development

ATLAS: How it works

Three basic steps



Input:

Praat TextGrid and .wav file

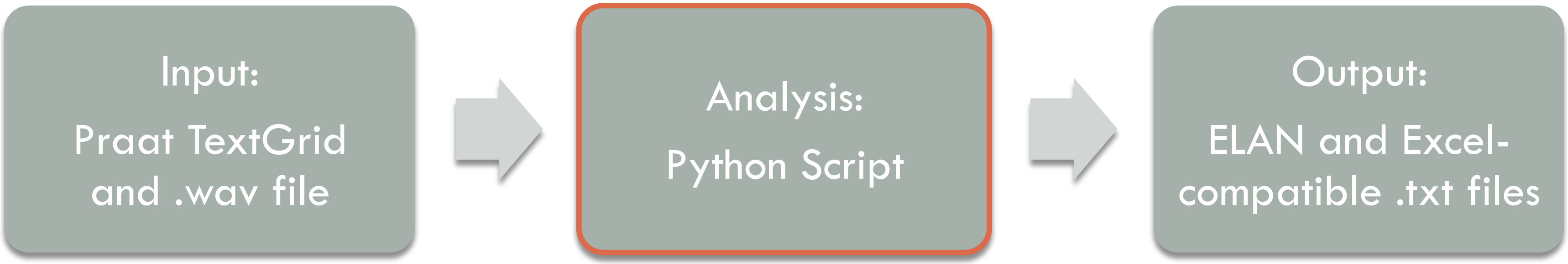
Analysis:

Python Script

Output:

ELAN and Excel-compatible .txt files

Three basic steps



Input:

Praat TextGrid and .wav file

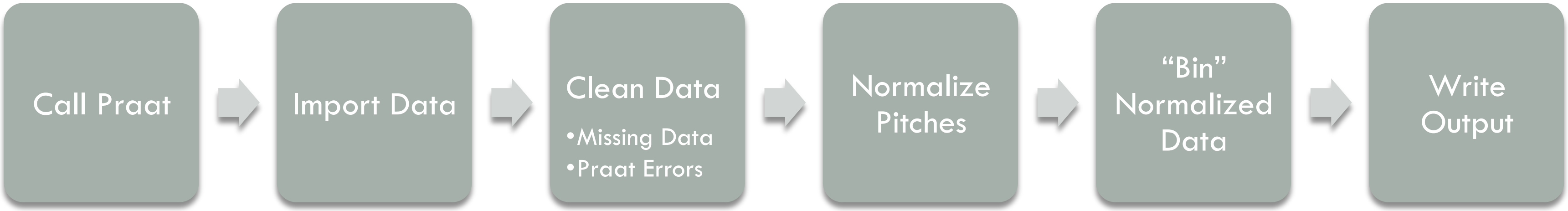
Analysis:

Python Script

Output:

ELAN and Excel-compatible .txt files

Python script: a closer look



Call Praat

Import Data

Clean Data

- Missing Data
- Praat Errors

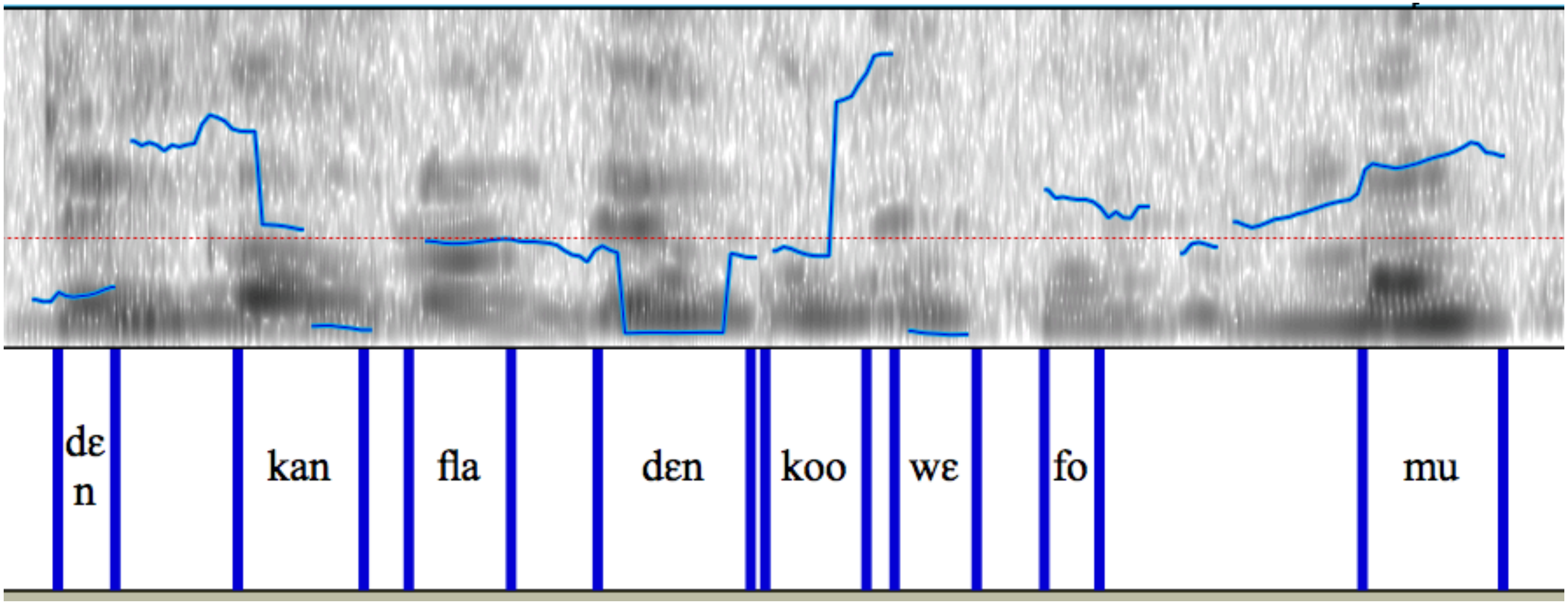
Normalize Pitches

“Bin” Normalized Data

Write Output

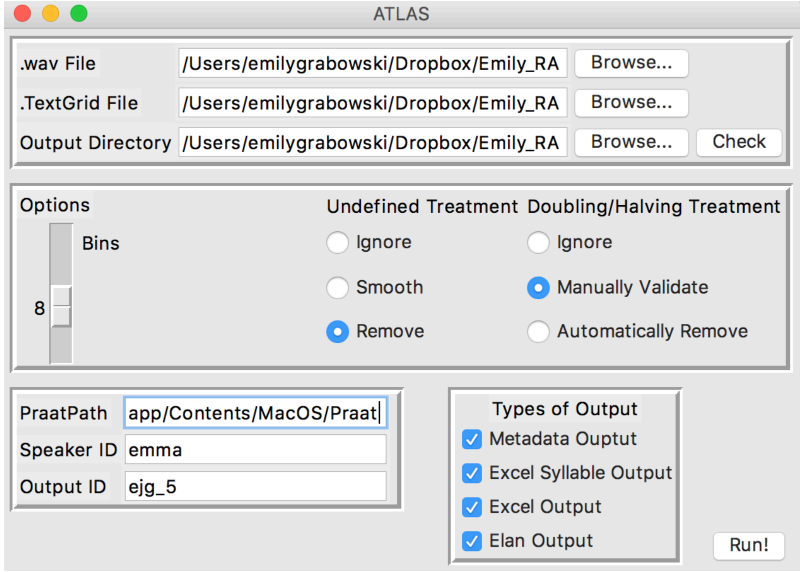
Input

- Semi-automated version requires a .wav file and Praat .TextGrid annotations as input
  - Annotate each TBU



Python script: a semi-automated tool

- Prompts the user to input arguments
- Currently command line tool
  - Initial arguments are input via a form
- Does not require the user to interface directly with Python



ATLAS

.wav File /Users/emilygrabowski/Dropbox/Emily_RA

.TextGrid File /Users/emilygrabowski/Dropbox/Emily_RA

Output Directory /Users/emilygrabowski/Dropbox/Emily_RA

Options

Bins 8

Undefined Treatment
○ Ignore
○ Smooth
● Remove

Doubling/Halving Treatment
○ Ignore
● Manually Validate
○ Automatically Remove

PraatPath app/Contents/MacOS/Praat

Speaker ID emma

Output ID ejg_5

Types of Output
☑ Metadata Ouptut
☑ Excel Syllable Output
☑ Excel Output
☑ Elan Output

Python Script: Cleaning the Data

Error	Cause(s)	How we deal with it
Undefined: Praat pitch-tracking not obtaining a signal	• Noisy/poor quality recording • TextGrid capturing part of a consonant	1. Remove all tokens with errors, or, 2. Smooth tokens with errors on boundaries, remove the rest
The Octave Problem (Doubling/Halving)	• Praat's algorithm thinks that the pitch is either an octave higher or lower • More often found with speakers with larger-than-average ranges	1. Automatically remove flagged tokens, or, 2. Manually confirm doubling/ halving
Outliers: Other	• Praat picked up data from outside the speaker • Speaker had one really high or low token	After correcting for the above two categories of errors, fit to a normal distribution (within 3 SDs) to find the speaker's probable range.

Python script: normalization

- Normalization allows for better comparison between speakers
- Hertz → semitones
  - cf. Baken 1987, Hart et al. 1990, Liberman and Pierrehumbert 1984, Ross et al. 1986, Xu 2004, etc.
  - A measure of frequency based on number of 'half-steps' (in the Western musical tradition) from a reference tone
  - Reference tone is the speaker's mean pitch in Hertz (after outlier correction)
  - Equation: $12(log_2(freq/ref))$

Python script: creating bins

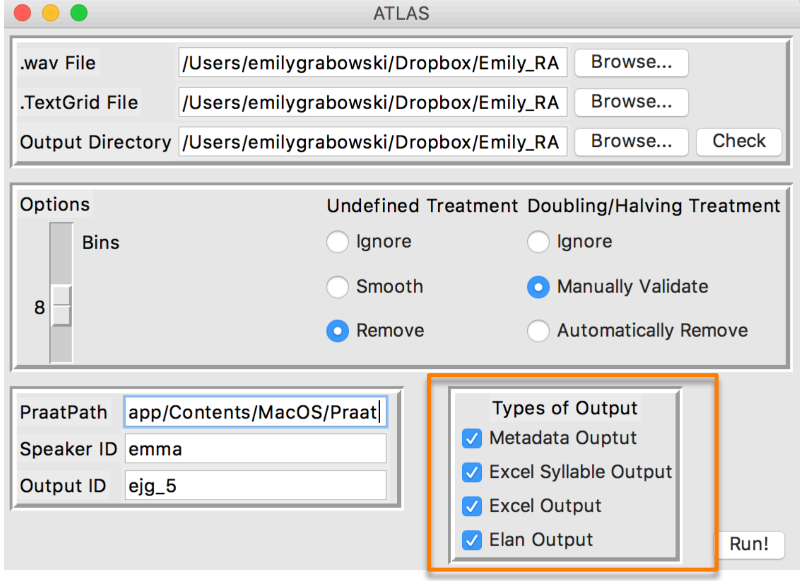
- Start with speaker's overall range (corrected for outliers)
- Range is divided up into equal parts (equal bins)
- User can specify the number of bins that they wish to use
  - More bins = more phonetic detail
  - We have found 8 to be a good number so far

Python script: assigning tokens to bins

- Take samples throughout the TBU
- Two extremes:
  - Could take as much as every 1/100th second throughout the TBU
    - Could be time-normalized for analysis
    - Can be overwhelming amount of detail
  - Could also do average for the overall token
    - Loses contour tone/phonetic detail
- Compromise: Measure at 20%/80%
  - Avoids consonant effects
  - Preserves contours and most important phonetic details

Output

- Desired output can be selected at the beginning:
- Main types
  1. ELAN-compatible (minimalist: time stamp + bins)
  2. Detail-rich spreadsheets
     - Two points (20%/80%) per syllable
     - Every 1/100s per syllable
  3. Metadata



ATLAS

.wav File /Users/emilygrabowski/Dropbox/Emily_RA
.TextGrid File /Users/emilygrabowski/Dropbox/Emily_RA
Output Directory /Users/emilygrabowski/Dropbox/Emily_RA

Options
Bins 8

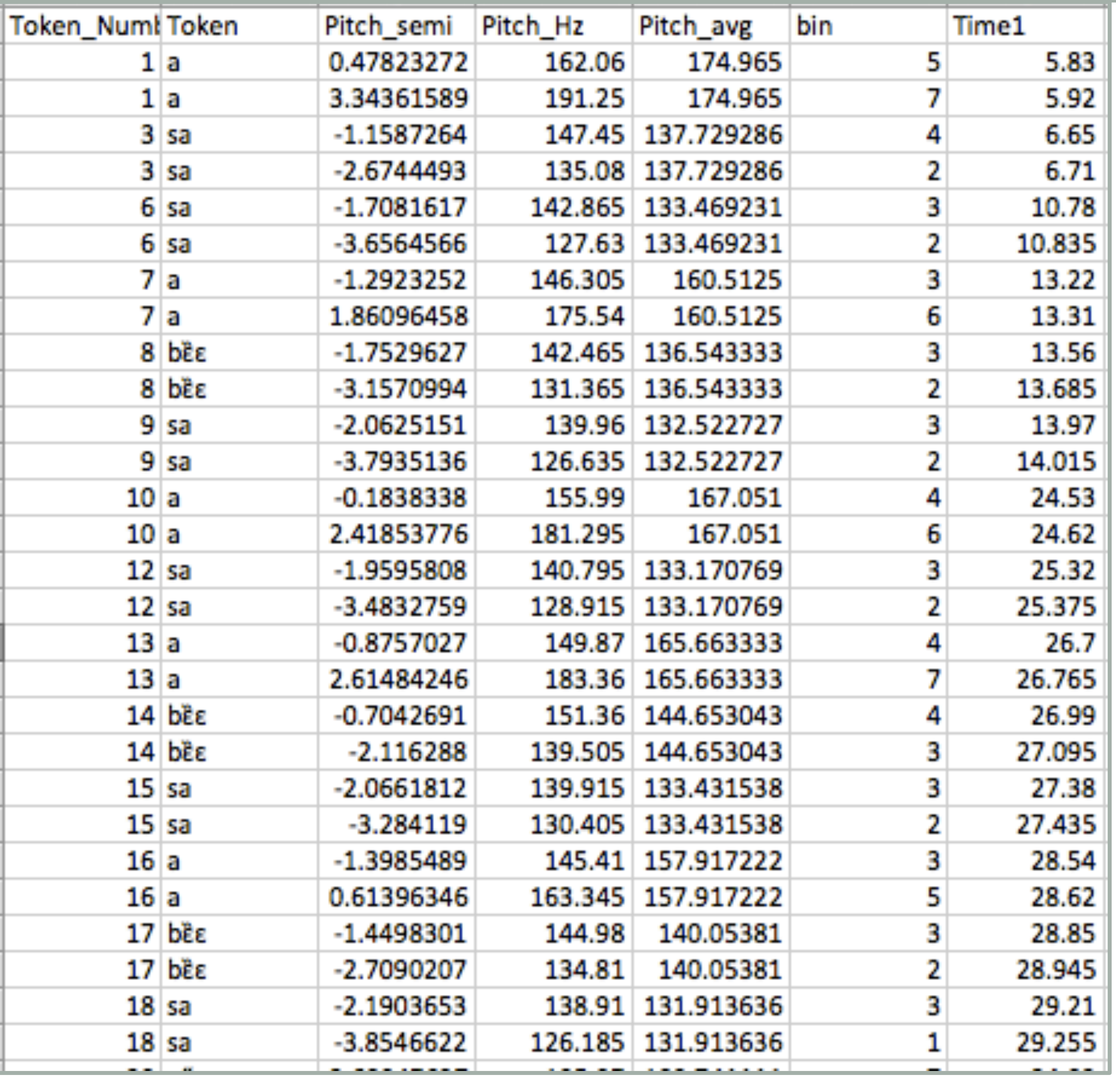
Undefined Treatment
○ Ignore
○ Smooth
● Remove

Doubling/Halving Treatment
○ Ignore
● Manually Validate
○ Automatically Remove

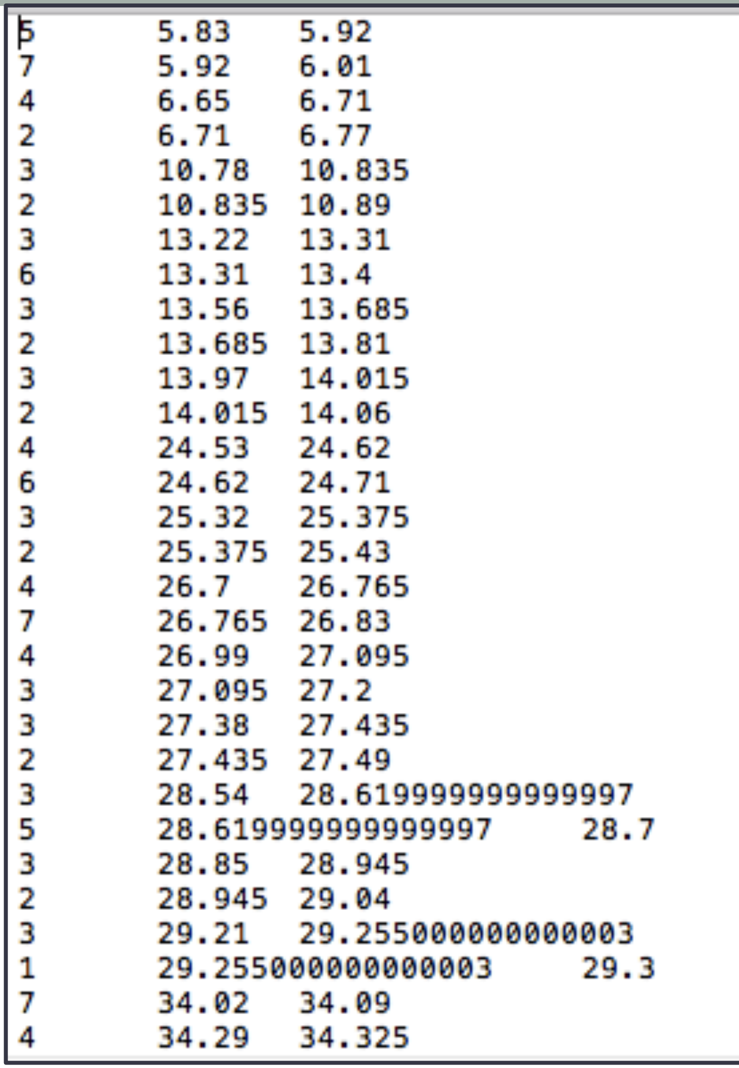
PraatPath app/Contents/MacOS/Praat
Speaker ID emma
Output ID ejg_5

Types of Output
☑ Metadata Ouptut
☑ Excel Syllable Output
☑ Excel Output
☑ Elan Output

Output: .txt files

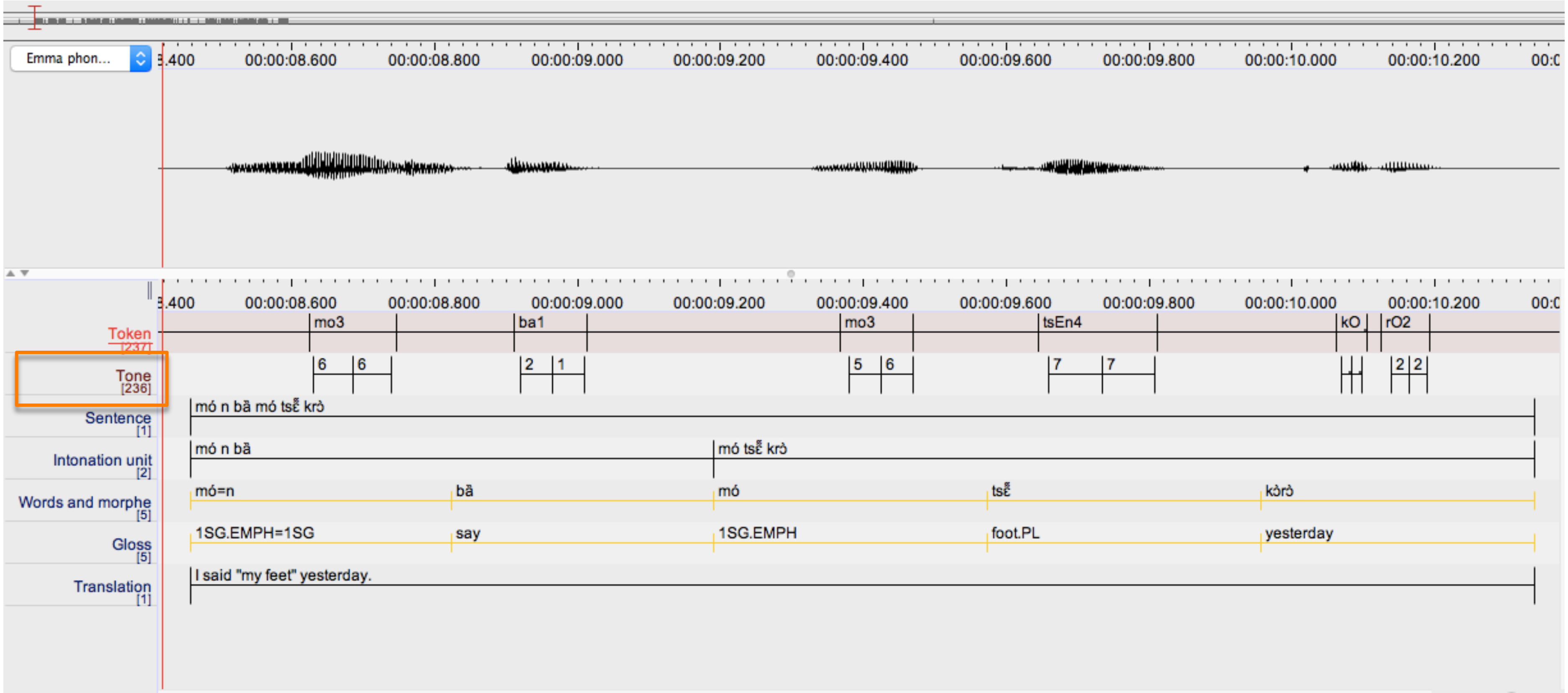


Token_Numl	Token	Pitch_semi	Pitch_Hz	Pitch_avg	bin	Time1
1	a	0.47823272	162.06	174.965	5	5.83
1	a	3.34361589	191.25	174.965	7	5.92
3	sa	-1.1587264	147.45	137.729286	4	6.65
3	sa	-2.6744493	135.08	137.729286	2	6.71
6	sa	-1.7081617	142.865	133.469231	3	10.78
6	sa	-3.6564566	127.63	133.469231	2	10.835
7	a	-1.2923252	146.305	160.5125	3	13.22
7	a	1.86096458	175.54	160.5125	6	13.31
8	bɛ̈ɛ	-1.7529627	142.465	136.543333	3	13.56
8	bɛ̈ɛ	-3.1570994	131.365	136.543333	2	13.685
9	sa	-2.0625151	139.96	132.522727	3	13.97
9	sa	-3.7935136	126.635	132.522727	2	14.015
10	a	-0.1838338	155.99	167.051	4	24.53
10	a	2.41853776	181.295	167.051	6	24.62
12	sa	-1.9595808	140.795	133.170769	3	25.32
12	sa	-3.4832759	128.915	133.170769	2	25.375
13	a	-0.8757027	149.87	165.663333	4	26.7
13	a	2.61484246	183.36	165.663333	7	26.765
14	bɛ̈ɛ	-0.7042691	151.36	144.653043	4	26.99
14	bɛ̈ɛ	-2.116288	139.505	144.653043	3	27.095
15	sa	-2.0661812	139.915	133.431538	3	27.38
15	sa	-3.284119	130.405	133.431538	2	27.435
16	a	-1.3985489	145.41	157.917222	3	28.54
16	a	0.61396346	163.345	157.917222	5	28.62
17	bɛ̈ɛ	-1.4498301	144.98	140.05381	3	28.85
17	bɛ̈ɛ	-2.7090207	134.81	140.05381	2	28.945
18	sa	-2.1903653	138.91	131.913636	3	29.21
18	sa	-3.8546622	126.185	131.913636	1	29.255



```
5       5.83    5.92
7       5.92    6.01
4       6.65    6.71
2       6.71    6.77
3       10.78   10.835
2       10.835  10.89
3       13.22   13.31
6       13.31   13.4
3       13.56   13.685
2       13.685  13.81
3       13.97   14.015
2       14.015  14.06
4       24.53   24.62
6       24.62   24.71
3       25.32   25.375
2       25.375  25.43
4       26.7    26.765
7       26.765  26.83
4       26.99   27.095
3       27.095  27.2
3       27.38   27.435
2       27.435  27.49
3       28.54   28.619999999999997
5       28.619999999999997      28.7
3       28.85   28.945
2       28.945  29.04
3       29.21   29.255000000000003
1       29.255000000000003      29.3
7       34.02   34.09
4       34.29   34.325
```

Output: tonal tier in ELAN



Token [237]: mo3 | ba1 | mo3 | tsEn4 | kO | rO2

Tone [236]: 6 | 6 | 2 | 1 | 5 | 6 | 7 | 7 | 2 | 2

Sentence [1]: mó n bă mó tsɛ̃ krò

Intonation unit [2]: mó n bă | mó tsɛ̃ krò

Words and morphe [5]: mó=n | bă | mó | tsɛ̃ | kòrò

Gloss [5]: 1SG.EMPH=1SG | say | 1SG.EMPH | foot.PL | yesterday

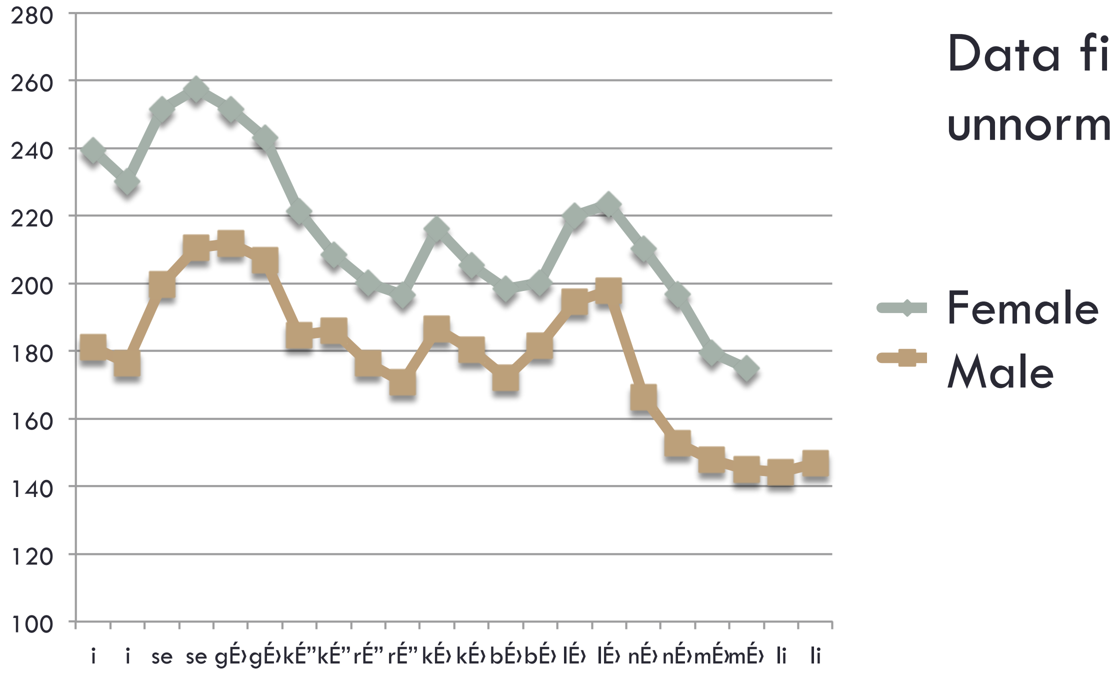
Translation [1]: I said "my feet" yesterday.

ATLAS: Research applications

Phonetic realization of tone

- With finer grain settings, phonetic realization can be visualized
- Case study: Tommo So (Dogon, Mali)
  - Two phonemic tones (H, L), plus surface underspecification (0)
- Controlled elicitation data from three speakers
  - 2 male, 1 female

Phonetic realization of tone



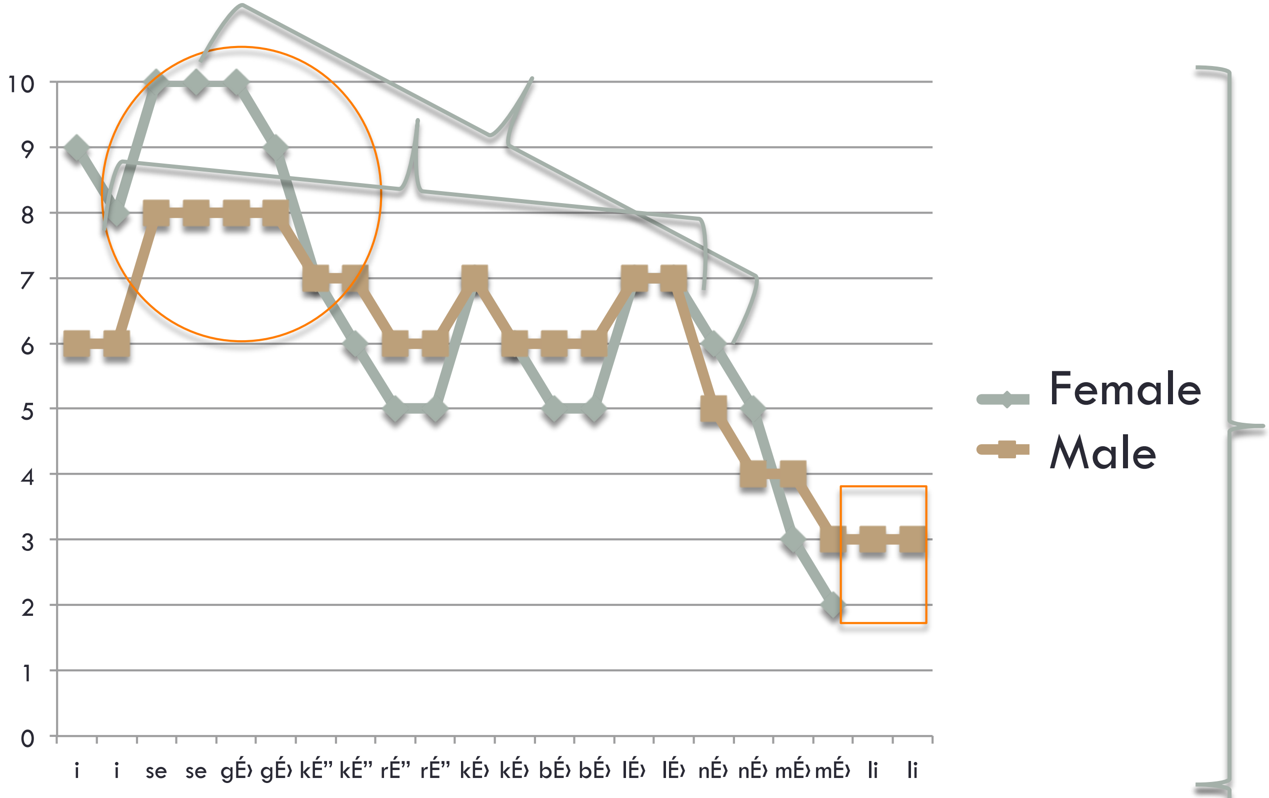
Data first in Hz (f0), unnormalized

L H 0 L L L L H L L H

ìsé = gɛ kɔ̀rɔ̀ kɛ̀bɛ̀lɛ́ nɛ̀mɛ̀-lí

‘the dog didn’t lick the calabash shard’

Phonetic realization of tone



ATLAS

L <u>H 0 L</u> L L L H L L H

ìsé = gɛ kɔ̀rɔ̀ kɛ̀bɛ̀lɛ́ nɛ̀mɛ̀-lí

‘the dog didn’t lick the calabash shard’

Human analysis

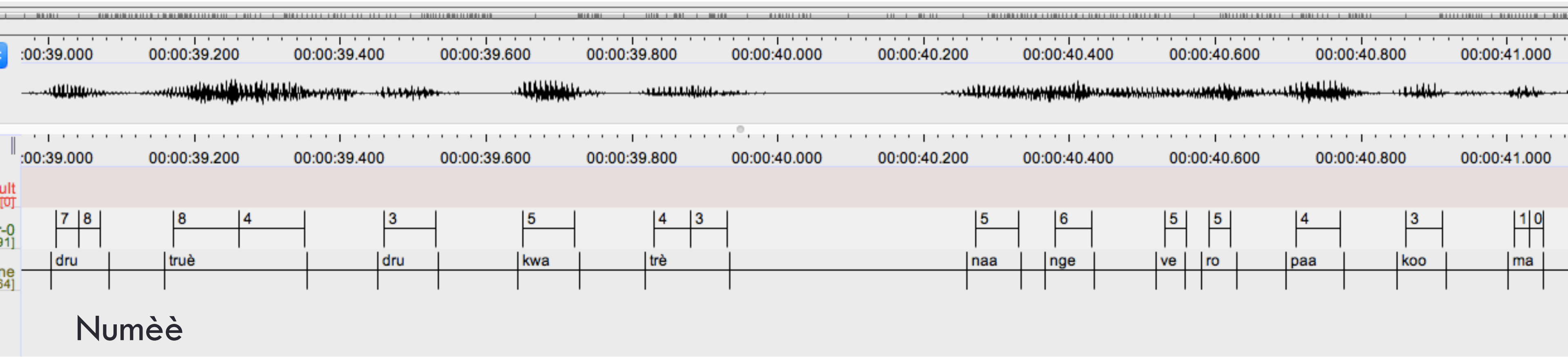
Confirming the literature

- Can be used in early stages of work to confirm descriptions in the literature
- Case study: Kwényï and Numèè (New Caledonia)
- Both languages are (probably) tonal, but neither tone system well understood
- Rivierre (1973) reports Numèè and Kwényï are mutually intelligible, but with opposite tone systems
  - Numèè overall falling melodies
  - Kwényï overall ascending ("plaintive") melodies

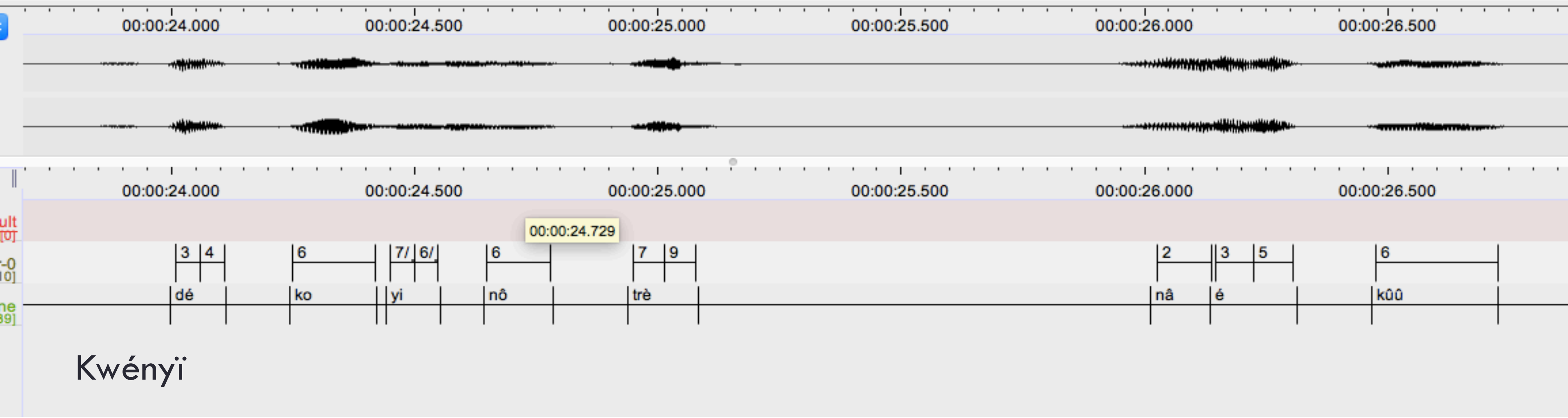
Confirming the literature

- Created TextGrids for a Kwényï narrative recorded in 2016 and a recording of Numèè from the LACITO archives
  - Both versions of a classic Melanesian “rat and the octopus” story
- Ran ATLAS with 10 bins

Confirming the literature



Numèè



Kwényï

Confirming the literature

- □ Average beginning and end levels for intonational units confirms the literature

	Average sentence beginning level	Average sentence ending level
Numèè	3.9	2
Kwényï	2.7	5.7

- □ Though Numèè typically rises before it falls
- □ Also an example of including tonal annotations for Kwényï before tonal analysis is complete

Conclusions

Summary of ATLAS

- Semi-automated tool to produce broad phonetic tone transcriptions
- User-friendly, requiring no programming knowledge and no prior experience with tone
- Transcriptions can be imported into ELAN

Summary of ATLAS

- ATLAS is **not** meant to:
  - Automate phonological analysis
  - Replace the need for phonological analysis and subsequent marking of tone
- Phonetic tonal annotations promote transparency and replicability
  - Whether alongside phonological analysis or on their own

Future development

- Fully automate, creating web and desktop versions
  - Forced alignment (e.g. FAVE, Rosenfelder et al. 2011)
  - Better interface with ELAN
- Optimization and development
  - Outliers
  - Doubling/halving
  - Maintaining speaker databases across recordings

To download the beta version…

- Go to dartmouth.edu/~mcpherson and follow the link on the home page.

Acknowledgments

- We would like to thank the Dartmouth College Neukom Institute and Office of Undergraduate Advising and Research for financial support. Many thanks to our colleagues, particularly Jim Stanford, for helpful discussions and feedback in the development of this prototype.

References

Baken, Ronald. J. (1987). *Clinical measurement of speech and voice*. Boston: College Hill Press.

Bird, Steven. 1994. Bird, S. (1994). Automated tone transcription. Retrieved from http://arxiv.org/abs/cmp-lg/9410022

Bird, Steven and Haejoong Lee. 2014. Computational support for early elicitation and classification of tone. *LDC* 8: 453-461.

Cooper-Leavitt, J. E. (2016). A computational classification of Thai lexical tones. *The Journal of the Acoustical Society of America*, *139*(4), 2216–2216.

Cruz, E., & Woodbury, A. C. (2014). *Finding a way into a family of tone languages: The story and methods of the Chatino language documentation project. Language Documentation & Conservation* (Vol. 8).

Hart, Johan T., Collier, René, & Cohen, Antonie. (1990). *A perceptual study of intonation: An experimental approach to speech melody*. Cambridge: Cambridge University Press.

Liberman, M. and J. Pierrehumbert (1984) Intonational Invariance under Changes in Pitch Range and Length , in M. Aronoff and R. Oehrle, eds, *Language Sound Structure*, MIT Press, Cambridge MA. 157-233.

McPherson, Laura. 2011. *Tonal underspecification and interpolation in Tommo So*. MA Thesis, UCLA

McPherson, Laura. 2013. *A Grammar of Tommo So*. Berlin: De Gruyter Mouton.

References

Rivierre, Jean-Claude. 1973. *Phonologie comparée des dialectes de l'extrême-sud de la Nouvelle Calédonie*. Paris: SELAF.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction) program suite. Software.

Ross, Elliott D., Edmondson, Jerold A., & Seibert, G. Burton. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics* 14(2):283–302.

Tan Lee, Ching, P. C., Chan, L. W., Cheng, Y. H., & Mak, B. (1995). Tone recognition of isolated Cantonese syllables. *IEEE Transactions on Speech and Audio Processing*, *3*(3), 204–209.

Wu, Jiang, Zahorian, Stephen A., & Hu, Hongbing. (2013). Tone Recognition for Continuous Accented Mandarin Chinese. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, 2013, pp. 7180-7183.

Xu, Yi (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics 5:757-97*.

Yang, W.-J., Lee, J.-C., Chang, Y.-C., & Wang, H.-C. (1988). Hidden Markov model for Mandarin lexical tone recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *36*(7), 988–992.

Data Organization

- Calls Praat script from Python (uses praatIO module developed by Tim Mahrt)
- Automatically imports the results
- TokenList:
  - Info (speaker id, etc.)
  - Token1
    - Info (e.g. # undefined tokens)
    - [Pitchentry1, Pitchentry2…]
  - Token2
    - Info
    - [Pitchentry1, Pitchentry2...]